Reviewer’s report

Title: The impact generated by public and charity funded research in the UK: A systematic literature review

Version: 0 Date: 21 Oct 2018

Reviewer: Stephen Hanney

Reviewer’s report:

HRPS-D-18-00204: The impact generated by public and charity funded research in the UK: A systematic literature review

Reviewer's comments

This review is highly relevant for Health Research Policy and Systems, and is generally well-written. It has potential for publication as it has identified a range of papers on the impact generated by public and charity funded health research in the UK published between 2006 and 2017, and made progress in collating and analysing the findings from them.

However, in order to contribute a useful critical analysis I believe a rather more precise, detailed and nuanced analysis is required that builds on some of the key points already included. In terms of precision, in several of the numbered points below I indicate how the accuracy of the paper might be improved. For example, by addressing the fact that two papers describing exactly the same study findings have been included, and, further, are given different scores from each other for the risk of both funding bias and selection bias. Additionally, it appears as if at least one included paper does not purport to be an assessment of impacts, but rather is an interesting analysis of research inputs. Some of the definitions of categories of risk look questionable, for example the one in relation to funding. This definition results in Sainty (2013) being the only paper categorised as low risk of bias for funding, and yet the author conducted the impact assessment as the paid R&D manager of the research programme whose impact she was assessing.

The current review also makes a harsh assessment that most of the papers have a 'high risk' of reporting bias and 'failed to provide supporting evidence when making claims'. In various comments below related to specific points in the text I highlight how this current review itself appears to have overlooked data that are provided in some of the studies. Such omissions, in turn, might raise some doubts about the accuracy of some of the scores given for the risk of reporting bias. The authors might feel, as they imply in the limitations in relation to the article by Morgan Jones et al, that they did not have the resources to tackle the challenge of extracting precise data from each of the studies. However, in that case, it is unclear how the authors of the current review could conclude that little evidence had been provided, and that, therefore, such papers should be classified as 'high risk' for reporting bias. My comments at various stages below should help the authors identify some of the missing, and/or inaccurate, data in their analysis.
Furthermore, to provide a more nuanced analysis it would be useful to consider issues such as the balance between the desire to provide a detailed analysis, as in some of the included papers, and the cost of doing so. The latter concern helped inform some of the new approaches to impact assessment. Furthermore, the review could usefully build on important points made at the start of the final paragraph of the Conclusion about research activity having a degree of risk and unpredictability, and the consequent implications for impact assessment. It might be helpful if these points were brought into the Background section and informed the analysis in various ways. For example, while the aims of each study are described in the Table, the implications of some of them do not seem to be fully taken into account in various parts of the analysis. If research activities are unpredictable, and if, as with the Morgan Jones et al (2016) study, the stated aim is to identify 'evidence of NIHR funded research which has generated benefits to and wider impacts on health', then it could be argued that sometimes a selective selection approach is the most appropriate one to use to achieve the aim. Comments about the need for a more nuanced approach are also incorporated in the list of numbered points below:

1. P.2 Abstract: might need to be revised following various revisions to the text.

2. P.3: Background: Various points about the nature of research, and the implications for impact assessment could usefully be brought in here, along with a recognition that the purposes of research impact assessments might vary and different categories of research are more amenable to impact assessment than others. It might also be useful to set the paper in the context of the existing literature to a greater extent, for example by drawing on the analysis in the reviews of the literature in the Hanney et al (2007) paper and Raftery et al (2016) - both of which are already referenced in the current review.

3. P.4, lines 30-35: it would be helpful to explain why the journals hand-searched were thought to be 'key' when they did not provide any of the papers included in the review, and only BMC Health Services Research provided any papers referenced at all in the study.

4. P4, Study selection: it would be helpful to explain the approach to duplicate publications from the same study. Hanney et al (2007) is a Health Technology Assessment report and such reports have a highly unusual, but significant role. They provide the opportunity for research teams to present extremely detailed accounts of the aims, methods, findings and implications of their studies. In many ways they are 'monographs', but because they undergo rigorous peer-review they have been accepted as articles in the journal Health Technology Assessment which was given an ISI Web of Knowledge Journal Impact Factor. However, because these are such detailed publications, the HTA Programme encourages authors also to publish much briefer versions of their papers in traditional journals. Therefore, in this instance, Raftery et al (2009) is a brief journal article on the same data as described in Hanney et al (2007).

5. P.6, Table 1: Further refinement is required of the categorisations for funding bias. The low risk category is defined as 'The study did not receive any funding'. However, the only study analysed as being in this category (Sainty et al) raises considerable doubts about the adequacy of this definition as it stands. Sainty herself points out the conflicts of interest in
her study, because she conducted it as part of her job as the R&D manager of the research fund whose projects she was assessing. So, it seems odd that those circumstances should lead to a 'low risk' categorisation, but studies that were conducted totally, or primarily, by research teams independent of the sponsor (albeit funded by the sponsor) should be categorised as being 'high risk', (including Bunn et al; Guthrie et al, 2015, and 2016; Hanney et al, 2017 and 2013; Lichten et al, 2017; Morgan Jones et al, 2016). Neither is it explained why a study receiving funding from a different funding body than the one it is evaluating should be categorised as 'medium risk' rather than 'low risk'. Similarly, is it appropriate that a study where the funding is not stated is viewed as being only 'medium risk', whereas studies that are transparent about their funding from the sponsor are 'high risk'? This also seems to contradict the approach taken on selection bias, where if the response rate is not reported, the study is viewed as being 'high risk'.

6. P.6, Study inclusion: compared to the review in Hanney et al (2007) this current study has missing data because it does not list all the papers (48) taken to full review. Inclusion of such a list would allow the informed reader to check, for example, whether Kuruvilla et al (2007) had been identified in the search, but excluded, or had not been identified (this is the paper that describes the application of the approach presented in Kuruvilla et al, 2006).

7. P6/7, Study characteristics: As noted above, I do not believe the study by Hall et al is a research impact study, and neither do I believe it is accurate to state that it looks at 'the effect of health research on tobacco and disease burden'. What it does do is make an important analysis of how far the funding for tobacco research matches the burden of disease associated with tobacco. Therefore, it is an important analysis of research inputs, not impact. As also noted, it might be useful to discuss the purpose of the assessment studies because of the implications that might have for the appropriateness of different methods. Furthermore, I think the statement that the design of the studies was such that none of the studies could establish causality between research funding and various outcomes might need some disaggregation. Will there not be differences in terms of clarity of causality, for example, between, on the one hand research funding and the production of articles, and on the other hand research funding and improved health?

P.8/9: Risk of bias and Table 2: it is good that the authors picked up that the Hanney et al (2007) paper had used a random selection approach, and thus was low risk for selection bias, but it is not clear why for the identical study (Raftery et al) was categorised as being medium risk. In addition to the points made earlier, much further analysis is required around the statement that 'many studies stated that impact was generated but provided little supporting evidence (9/15 or 60%).' Presumably this refers to the 9 studies put in the high risk category for reporting bias. A proper analysis of the Guthrie et al (2015) paper, for example, would reveal that this was published as a report in the Health Technology Assessment series. It is over 300 pages long and contains an enormous amount of supporting evidence, as noted in the following extract from their paper (Guthrie et al, 2015, p.xxi):
We explored a wide range of impacts resulting from HTA programme-funded research and the HTA programme. We carried out an analysis of impact across the HTA programme using the following methods:

* Interviews (n=20) Senior stakeholders from academia, policy-making organisations and the HTA programme.

* Bibliometric analysis Citation analysis of publications (n=1087) arising from HTA programme-funded research.

* Researchfish survey Electronic survey of all HTA grant holders (n=619) [excluding Technology Assessment Reports (TARs)].

* Payback case studies (n=12) In-depth case studies of HTA programme-funded research, which included document review, interviews and bibliometric analysis.

This multi-method study allowed us to synthesise data from multiple sources to identify key findings regarding the impact of the HTA programme.'

It should be noted that these data include an analysis of all the data provided by the 619 relevant projects in their submission to Researchfish (ie, a method which is promoted in the Conclusion of the current review as a way to provide a more systematic data collection). But, the Guthrie et al study goes much further because they have also supplemented the Researchfish data by adding a major citation analysis exercise and 12 extremely detailed case studies. I find it difficult to see how it could be correct to conclude that the Guthrie et al study 'provided little supporting evidence'. Similarly, the HTA report by Hanney et al (2007) is 200 pages long and contains a full analysis of the evidence from a comprehensive survey and 16 detailed case studies, each of which uses the elements of the Payback Framework to report the evidence from an interview, extensive documentary analysis and triangulation. Furthermore, various other papers report considerable evidence from extensive, and often diverse, data collection activities. These studies include: Hanney et al, 2013; Lichten et al, 2017; Morgan Jones et al, 2016; Peckham et al, 2008.

8. P.9, line 46: after stating the average number of publications reported by Hanney et al (2007) for the HTA projects, the current paper goes on to state: 'Raftery and colleagues [15] confirmed the results and, in addition, showed the mean number of conference presentations was 5.2 per project.' As noted, there is a question as to whether Raftery et al, as a duplicate study, should be included at all, but, irrespective of that, this text seems to imply that Raftery et al provided additional information in relation to the number of conference papers that was not in Hanney et al. In fact, the opposite was the case with Hanney et al not only providing the average 5.2 per project figure, but also the figure for each of the 3 types of projects included in the analysis that came to an overall average of 5.2.

9. P.11, line 55: the final paragraph refers to 'five' of the reviewed studies providing a series of examples through which funded research has contributed to the development of research capacity. It misses out some further examples. For example, while it was not the
most important payback category in the Hanney et al (2007) study, p.55 provides the total number of qualifications gained according to the survey, and each of the 16 case studies has capacity building etc as a heading, and reports whether or not, and how, it arose in the particular case.

10. P.14, top line: it is misleading to suggest just 2 studies used a survey to identify impacts on informing policy. Additionally, at least 2 more studies drew on survey evidence as part of a mixed-method approach to assessing the impact on policy, namely Guthrie et al (2015), who used the Researchfish survey data, and Hanney et al (2007).

11. P.15, top paragraph: the text correctly notes the important role for research in informing policy that was recorded in the assessments by Guthrie et al and Hanney et al. The text also correctly notes some of types of policy informed by HTA research. In order fully to correct the point made earlier in the text by the authors (and repeated later) that limited supporting evidence was provided, it would be worth the authors of the current text going further. They could appropriately note that the case study assessments on the impact of the HTA programme went into considerable detail about the exact names of the policy documents informed by specific HTA projects, and the precise issues in the documents that were influenced by the specific research.

12. P.15, middle para, last line: I don't think the meaning of this sentence is clear.

13. P.15, last sentence: again, to correct the statement about limited supporting evidence, it might be useful to note that Hanney et al (2013) also provided detailed examples of research making an impact on health in several of the case studies.

14. P.16: middle para: I do not think the Hall et al study meets the inclusion criteria.

15. P.17: several further studies also addressed the question of health gains. These include both Glover et al (2014) and Guthrie et al (2016) for whom the analysis of the rate of return is based on first establishing the level of health gain from research. Each of the 16 case studies in Hanney et al (2007) considered the question of health gain, even in those instances where it was to report that there was no evidence.

16. P.19/20: Conclusions:

A) The Conclusion raises various important points, but in general needs to go much further in providing a nuanced analysis of the issues around the assessment of the impact of health research.

B) The review is correct to open up the issues around sources of bias. But, as noted, some of the account, especially around reporting bias, would benefit from considerable further analysis. Such analysis might give recognition to the extensive triangulation and presentation of evidence (especially around impact on policy) that is contained in some of the papers that have been described as failing ‘to provide supporting evidence when making claims.’ It is through the detailed
case studies that issues such as attribution are addressed, and detailed evidence provided about the level of impact that can justifiably be claimed from specific pieces of research.

C) One issue that could usefully be discussed is the balance between the desirability of providing the detailed evidence in case studies, such as those in Hanney et al (2007) and Guthrie et al (2015), and the resources required to do this. Aspects of this are already discussed in several of the included studies and the wider literature referenced in the text. This crucial point in the existing literature feeds into thoughts about new approaches.

D) The authors are right to explore new approaches, but need to analyse them in the light of the findings of the review. So, Researchfish does provide a way of recording data on a large scale, but it is entirely self-reported and does not include the triangulation involved in the detailed case studies. It is thus open to much larger concerns about aspects of bias than arise with the case studies included in some of the papers included in the review. Similarly, the REF provides an important additional way to assess the impact of health research, but it is highly selective.

E) The issue of selectivity could usefully be considered in relation to the points made at the start of the final paragraph about the degree of risk and unpredictability of research, and that impact assessment can be interpreted as neglecting the inherent value of science. Some of the assessments of the impact of research that have adopted a selective approach are partially attempting to reconcile the nature of science with the demands for evaluation. They are sometimes doing this by focusing on some of the projects that have achieved impact, but not subjecting those that have not made an impact to detailed (and potentially unnecessary?) analysis. Thus, depending on the purposes of the exercise, a more nuanced and selective approach might be both more feasible and justified in the particular circumstances of research impact assessment. Similarly, the studies by Glover et al (2014) and Guthrie et al (2016) recognise that not all research will necessarily make an impact, but take into account the costs of the entire body of research they are valuing and set the benefits from some of the research against that. In this way they are being conservative and reducing the risk that would have arisen had they compared the benefits from selective studies against the costs of just those specific studies, rather than against the costs of the whole research portfolio.

F) Limitations: as noted, the authors are right to note the challenges involved in extracting precise data from all the studies. However, they should go on to recognise that in some cases these challenges arise from the very depth of evidence provided in some of the papers. This rather contradicts the authors' repeated claims about insufficient evidence being provided in some of the papers, and yet it was such claims which, in turn, had been used as the basis for the authors' categorisation of the papers as having a high risk of reporting bias. A further limitation the authors should note is that by only considering the period
from 2006-17 they missed out quite a few UK peer-reviewed impact assessment studies that had been included in the 2007 review by Hanney et al. The authors appear to try to justify this period by claiming it coincides with the creation of the NIHR, but I do not see the relevance of that as this is not a study that focuses specifically on the research funded by the NIHR.

18., p.24: References: these are generally accurate, but the journal title seems to be missing from reference 16, and the publisher and location seem to be missing from reference 18.

**Level of interest**
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

It could be argued I have a non-financial competing interest in relation to this review as it includes, and scores the level of risk of bias, of 3 papers I co-authored, and another for which I served on the Advisory Committee.
I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal